

Modelling fish habitat preference with a genetic algorithm-optimized Takagi-Sugeno model based on pairwise comparisons

Shinji Fukuda, Willem Waegeman, Ans Mouton and Bernard De Baets

Abstract Species-environment relationships are used for evaluating the current status of target species and the potential impact of natural or anthropogenic changes of their habitat. Recent researches reported that the results are strongly affected by the quality of a data set used. The present study attempted to apply pairwise comparison to modelling fish habitat preference with Takagi-Sugeno-type fuzzy habitat preference models (FHPMs) optimized by a genetic algorithm (GA). The model was compared with the result obtained from the FHPM optimized based on mean squared error (MSE). Three independent data sets were used for training and testing of these models. The FHPMs based on pairwise comparison produced variable habitat preference curves from 20 different initial conditions in the GA. This could be partially ascribed to the optimization process and the regulations assigned. This case study demonstrates applicability and limitations of pairwise comparison-based optimization in an FHPM. Future research should focus on more flexible learning process to make a good use of the advantages of pairwise comparisons.

Shinji Fukuda
Kyushu University, 6-10-1 Hakozaki, Fukuoka 812-8581, Japan, e-mail: shinji-fkd@agr.kyushu-u.ac.jp

Willem Waegeman
Ghent University, Coupure links 653, 9000 Ghent, Belgium e-mail: Willem.Waegeman@ugent.be

Ans Mouton
Research Institute for Nature and Forest (INBO), Kliniekstraat 25, 1070 Brussels, Belgium e-mail: Ans.Mouton@inbo.be

Bernard De Baets
Ghent University, Coupure links 653, 9000 Ghent, Belgium e-mail: Bernard.DeBaets@ugent.be

1 Introduction

Ecological models are abstractions of natural systems and tools for understanding complex processes and mechanisms involved. Habitat preference models are used in order to extract habitat preference information of a target species from observation data, or to express expert knowledge on the species. Species-environment relationships can thus be quantified and used in the decision making on plans and management options for a target ecosystem. In practice, there are two types of habitat models: univariate and multivariate models. Univariate models have been widely applied, for which the main idea consists of using a set of univariate preference functions to represent possible habitat preference of a target species. These models are used for the assessment of current status and future impacts of habitat changes in both time and space [3]. A variety of methods has been proposed and employed in practical applications, which include the habitat suitability index [3], resource selection functions [11], and other specific models such as a genetic Takagi-Sugeno fuzzy model [6, 9]. In recent years, multivariate approaches have gained more popularity as computational systems have become powerful and freely available. These include machine learning methods [13], fuzzy rule-based systems [1, 15], and statistical regression tools [2, 8]. In habitat modelling, it is often reported that the results are affected by the quality and quantity of a data set used [5, 12], which is partly because of the uncertainties inherent to observation data. The development of a sound methodology to cope with the different quality of data contributes the better understanding and reliable assessment of target ecosystems.

Preference modelling has been one of the key topics in information sciences. Preference is used for ranking items and the ranking can be used for decision making. Recently, pairwise comparison has been gaining interests in this field, and is reported to be a sound methodology in preference learning [10]. Despite the intensive works in theory, the pairwise comparison has not yet been applied to preference modelling in ecology, and it seems to be a good approach to cope with the data with uncertainties such as observation errors. It would, therefore, be interesting to exemplify the applicability and limitations of an pairwise comparison-based approach.

Our aim is to apply pairwise comparisons to the optimization of a Takagi-Sugeno-type fuzzy habitat preference model (FHPM) using a genetic algorithm (GA [7]). The results were compared with the previously developed FHPM optimized based on mean squared error (MSE). Three independent data sets were used for training and testing of the models. This first application of pairwise comparison provides useful information for the development of reliable habitat assessment approach using observation data with uncertainty.

2 Methods

2.1 Data Collection

A series of field surveys focusing on Japanese medaka (*Oryzias latipes*) was carried out in an agricultural canal in Kurume City, Fukuoka, Japan. Field surveys were conducted on three sunny days: 14 October and 5 and 9 November 2004. Two study reaches were established in the same canal: a 50-m-long study reach (1.6–2.0 m in width, 0.3% gradient) was surveyed on 14 October and 5 November, and a 30-m-long study reach (0.8–1.4 m in width, 0.3% gradient) was surveyed on 9 November. Habitat use by Japanese medaka and four physical habitat characteristics—water depth (cm, henceforth referred to as depth), current velocity (cm s⁻¹, velocity), lateral cover ratio (% cover), and vegetation coverage (% vegetation)—in the study reach were surveyed. The study reach was first mapped, then habitat use by the fish was observed, and finally the physical habitat characteristics within the reach were measured.

In the following analyses, fish distribution data are expressed as the log-transformed observed fish population density in the i^{th} water unit ($FPD_{o,i}$; individuals per square metre), where the subscript o indicates observed, $i = 1, 2, \dots, N$, denotes the water unit, and N is the total number of water units (Table 1). The size of data sets ($=N$) is 139 for the data set of 14 October, 130 for that of 5 November, and 86 for that of 9 November, all of which contain vectors of four habitat variables (depth, velocity, cover and vegetation) and fish population density.

Table 1 Species distribution data along with the four habitat variables of depth, velocity, cover and vegetation, each of which was observed on 14 October, 5 November, and 9 November 2004, respectively.

Date		D	V	C	VEG	log ₁₀ (FPD+1)	presence	absence	prevalence
14 Oct.	maximum	65.0	61.3	50.0	100.0	1.64			
	mean	13.6	13.7	19.1	49.7	0.37	77	62	55.4
	minimum	2.0	1.5	0.0	0.0	0.00			
	SD	8.0	9.3	14.0	29.6	0.41			
5 Nov.	maximum	57.0	31.6	50.0	100.0	1.56			
	mean	13.6	8.8	20.6	59.5	0.35	71	59	54.6
	minimum	3.0	1.5	0.0	0.0	0.00			
	SD	8.4	5.6	14.8	31.2	0.40			
9 Nov.	maximum	32.0	44.7	50.0	100.0	1.88			
	mean	15.6	12.5	19.8	55.2	0.42	31	55	36.0
	minimum	5.0	1.5	0.0	0.0	0.00			
	SD	6.2	9.6	13.9	33.3	0.60			

SD, standard deviation; D, depth (cm); V, velocity (cm s⁻¹); C, cover (%);

VEG, vegetation (%);

FPD, observed fish population density (individuals per square metre)

2.2 Fuzzy Habitat Preference Model

A fuzzy habitat preference model (FHPM) was employed for describing the habitat preference of the target fish [9]. An FHPM is a 0-order Takagi-Sugeno model [14] that relates habitat variables to habitat preference by considering uncertainties such as fish behaviour and measurement errors of the habitat variables. Simultaneous optimization of all model parameters enables an FHPM to evaluate habitat preference in an interpretable way despite nonlinear, complex interactions between habitat variables and habitat preference.

The FHPM used here is a single-input single-output fuzzy system with four sets of if-then rules: two sets of four rules each for depth and velocity, and two sets of three rules each for cover and vegetation (14 rules in total). The input values (the four habitat variables) were expressed as symmetric triangular fuzzy numbers with centre α_m and spread σ_m , where m denotes the habitat conditions of depth, velocity, cover, and vegetation, in order to take into account the measurement errors and the spatial variance of the habitat variables. The observed value was used as the centre value α_m , and the spread σ_m was determined from the allowed variance when the reach was divided into water units: the spread of depth σ_d was 1 cm, that of velocity σ_v was 1 cm s⁻¹, and those of cover σ_c and vegetation σ_{veg} were 10% each, where the subscripts d, v, c and veg denote depth, velocity, cover and vegetation, respectively. Next, all input values were transformed by membership functions (Fig. 1) into membership values ranging from zero to unity. The membership functions for depth and velocity were defined according to the ecological characteristics of Japanese medaka (Figs. 1a and 1b), and those for cover and vegetation were defined to include the available range of each of these habitat variables (Figs. 1c and 1d). A uniform partition with triangular membership functions was used, but the first and last membership functions were allowed to be trapezoidal. Each membership function has a corresponding singleton value in the consequent part. The habitat preference was calculated by taking the weighted mean of the singleton values in the consequent part, using the membership degrees as weights. Singleton values in the consequent part in the range [0, 1] were determined by a binary GA so as to minimize an objective function described below. The present GA consisted of a population of 100 individuals, each of which has a 56-bit string (4 bits \times 14 singletons). The shape of the habitat preference curves (HPCs), i.e., a set of singletons for a habitat variable, was constrained to have a unimodal or monotone form with maximum preference of one. Specifically, the model was penalized by giving large values with regard to the objective function when the above regulation was violated. Based on the objective function employed, the GA repeatedly modified the model structure (the singletons in the consequent part) using three genetic operations: roulette wheel selection with an elitist strategy, uniform crossover and mutation at a probability of 5%. The optimal model was obtained after 2,000 iterations. The GA optimization was repeated using 20 different sets of initial conditions in order to evaluate the variance of the model structure that resulted from the initial conditions.

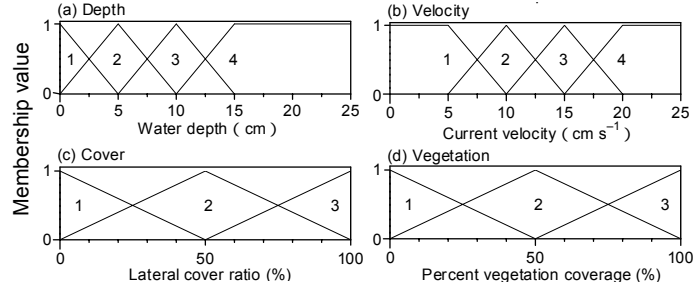


Fig. 1 Membership functions

2.3 Objective Function

We employed either of mean squared error (MSE) between observed and predicted fish population density or mean penalty (MP) evaluated from pairwise comparison as our objective function. The procedures to calculate each objective function are as follows.

First, the composite habitat preference according to the four habitat variables was calculated as

$$HP_i = P_{d,i} \times P_{v,i} \times P_{c,i} \times P_{veg,i} \quad (1)$$

where HP_i denotes the habitat preference in the i^{th} water unit, and $P_{d,i}$, $P_{v,i}$, $P_{c,i}$, and $P_{veg,i}$ are the habitat preference with respect to the individual habitat variables depth, velocity, cover and vegetation, respectively. Here, the use of multiplication in Eq. (1) produces an FHPM with higher performance and consistent habitat preference curves (HPCs) [6].

To calculate the MSE, fish population density of the Japanese medaka was estimated with Eq. (2):

$$FPD_{c,i} = \left(\frac{HP_i}{\sum_{i=1}^N HP_i} \right) \times \sum_{i=1}^N FPD_{o,i} \quad (2)$$

where $FPD_{c,i}$ is the calculated fish population density in the i^{th} water unit, and $FPD_{o,i}$ is the observed fish population density. The MSE is then calculated from the observed and predicted fish population density.

The MP is an average of penalty values assigned according to a set of pairwise comparison of habitat preference and observed fish population density of two different data points. That is,

$$MP = \frac{2}{N(N-1)} \sum_{\lambda_j \neq \lambda_k} \text{penalty}(\lambda_j, \lambda_k) \quad (3)$$

where λ_j and λ_k are two data points j and k . The pairwise comparison was performed as follows.

1. Calculate the composite habitat preference of data point j (HP_j) using an FHPM and Eq. (1).
2. For each of data points j and k ,
 - 2.1 Compare the habitat preference values (HP_j and HP_k), from which either of the relationships $HP_j > HP_k$, $HP_j = HP_k$, or $HP_j < HP_k$ is obtained.
 - 2.2 Compare the observed fish population density ($FPD_{o,j}$ and $FPD_{o,k}$), from which either of the relationships $FPD_{o,j} > FPD_{o,k}$, $FPD_{o,j} = FPD_{o,k}$, or $FPD_{o,j} < FPD_{o,k}$ is obtained.
 - 2.3 Compute a penalty based on the relationship obtained from 2.1 and 2.2.
 - a. If the relationships are the same between the habitat preference and fish population density, then no penalty is assigned.
 - b. If the relationships are the opposite between the habitat preference and fish population density, then assign a penalty of 1.
 - c. If either of the relationships is equal and the absolute difference of other relationship is smaller than a predefined value¹, then assign a penalty of 0.5. Otherwise, assign a penalty of 1.
3. Compute the MP by means of Eq. (3).

2.4 Model Application and Analyses

To illustrate the difference between the FHPMs obtained from two different objective functions, these models were compared in terms of model performance and habitat preference information retrieved from them. The first data set was used for model development (training), and the remaining two data sets were used for model evaluation (testing). That is, 20 FHPMs were developed from each data set, each of which was tested using two different data sets. From the 20 initial conditions used in the model development, the variance of the model structures was quantified by using performance measures and the HPCs. The model performance was evaluated by the MSE, the MP and the area under receiver operating characteristics curve (AUC). The AUC is often used when evaluating species distribution models for presence-absence data [6, 13] and it is independent of the objective functions in this study. The MSE and MP were calculated using the fish population density and the AUC was calculated after converting the fish population density into presence-absence data. Of these performance measures, the mean and standard deviations of the 20 FHPMs from different initial conditions were used as a measure of the predictive accuracy and of the variance of model structures, respectively. In an FHPM,

¹ Allowed absolute difference for fish population density was set at 0.1 and that for habitat preference was at 0.05 (about 5% of their entire range).

an HPC can easily be obtained by providing consecutive values in the range of the corresponding membership functions (in steps of 0.1) to the FHPM and plotting the output values against the habitat variables. Specifically, the HPC shape indicates the habitat preference information retrieved from the data set used.

3 Results

The two different models (FHPMs based on pairwise comparison and FHPMs based on MSE) produced HPCs with similar trends in preference (Figs. 2–3). The HPCs of the best models have almost the same shape with slightly different degrees of preference (solid lines in Figs. 2–3). We can, however, observe slight differences in the shapes between the two models at the water depth around 15–20 cm in the data sets of 14 October and 5 November. Slightly contradicting shapes among the 20 HPCs are found from the FHPMs based on pairwise comparison for depth in the data set of 5 November (Fig. 2(ii-a)) and for vegetation in the data set of 9 November (Fig. 2(iii-d)). In other words, some HPCs show monotone forms, but others show unimodal forms. The both FHPMs produced variable HPCs but at different habitat variables (Figs. 2–3). For instance, the HPCs obtained from the FHPMs based on pairwise comparison show a variance in depth and vegetation (Fig. 2), whereas the HPCs from the FHPMs based on MSE show a variance in cover (Fig. 3). This could be partially resulted from different penalization approaches in training the models.

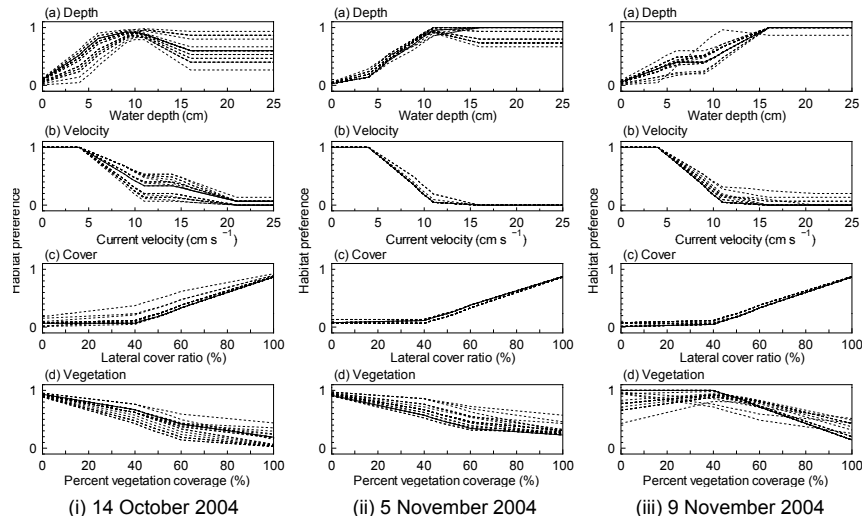


Fig. 2 Habitat preference curves obtained from fuzzy habitat preference model optimized based on pairwise comparison. Solid lines are the best curve with respect to the pairwise comparison and dotted lines are all curves obtained from 20 different initial conditions in a genetic algorithm

Performance of the two models differs by the performance measures (Table 2). The variance in HPCs seems to have some relationship with the standard deviation of all performance measures, except for the MP of FHPMs based on pairwise comparison. The FHPMs based on pairwise comparison show better performance in MP, while the FHPMs based on MSE show better performance in MSE and in AUC. The best FHPMs (with respect to each of the objective functions) from each data set also support the better performance of the FHPM based on MSE (Table 3). It can be seen from these results that FHPMs based on pairwise comparison failed to converge or reach at an optimal solution.

We observe no correlation between the MP and the AUC or the MSE (Fig. 4), although there could be some correlation between these measures within a data set. Despite the differences between the data sets, the MP of FHPMs based on pairwise comparison shows a small variance within the same data set, while the AUC and MSE plots spread vertically (Fig. 4(i)). This means that AUC and MSE cannot be minimized by minimizing the MP. The MSE of FHPMs based on MSE shows a small variance, while the MP and AUC show more variance (Fig. 4(ii)). The variance is dependent on the data set used as it can also be seen in Table 2. In contrast to the MP, the AUC and MSE show a positive relationship for both of the FHPMs (Fig. 5), of which the variance is small in the FHPMs based on MSE. This indicates that the minimization of MSE could lead to maximization of AUC through the GA optimization applied in this study.

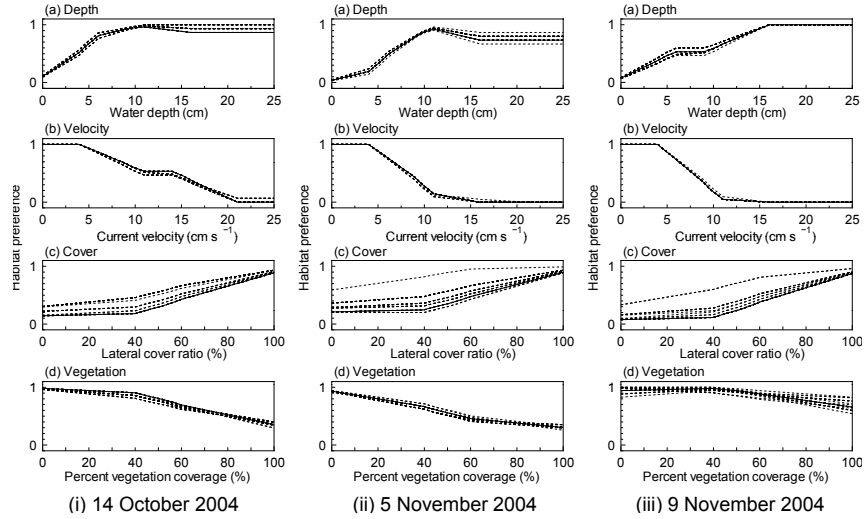


Fig. 3 Habitat preference curves obtained from fuzzy habitat preference model optimized based on mean squared error (MSE). Solid lines are the best curve with respect to the MSE and dotted lines are all curves obtained from 20 different initial conditions in a genetic algorithm

Table 2 Results of performance evaluation of fuzzy habitat preference models (FHPMs) optimized based on pairwise comparison (i) and FHPMs optimized based on mean squared error (MSE) (ii), of which model 1 is the FHPMs developed using the data set of 14 October, model 2 is those of 5 November, model 3 is those of 9 November 2004. Each model was tested using three data sets.

(i) pairwise comparison-based optimization				(ii) MSE-based optimization			
Data set used				Data set used			
MSE	Oct. 14	Nov. 5	Nov. 9	MSE	Oct. 14	Nov. 5	Nov. 9
Model 1	0.305	0.172	0.441	Model 1	0.069	0.084	0.216
	± 0.279	± 0.124	± 0.219		± 0.001	± 0.002	± 0.006
Model 2	0.371	0.123	0.224	Model 2	0.146	0.067	0.206
	± 0.044	± 0.021	± 0.035		± 0.007	± 0.000	± 0.009
Model 3	1.131	0.565	0.221	Model 3	0.240	0.115	0.143
	± 0.410	± 0.226	± 0.022		± 0.042	± 0.016	± 0.001
AUC	Oct. 14	Nov. 5	Nov. 9	AUC	Oct. 14	Nov. 5	Nov. 9
Model 1	0.738	0.735	0.747	Model 1	0.875	0.818	0.893
	± 0.100	± 0.080	± 0.100		± 0.002	± 0.007	± 0.011
Model 2	0.701	0.817	0.892	Model 2	0.786	0.860	0.912
	± 0.008	± 0.007	± 0.012		± 0.011	± 0.003	± 0.005
Model 3	0.623	0.680	0.792	Model 3	0.748	0.855	0.921
	± 0.028	± 0.066	± 0.050		± 0.012	± 0.003	± 0.003
MP	Oct. 14	Nov. 5	Nov. 9	MP	Oct. 14	Nov. 5	Nov. 9
Model 1	0.229	0.289	0.317	Model 1	0.277	0.321	0.406
	± 0.001	± 0.004	± 0.009		± 0.014	± 0.018	± 0.027
Model 2	0.339	0.236	0.235	Model 2	0.266	0.272	0.325
	± 0.001	± 0.001	± 0.007		± 0.009	± 0.012	± 0.017
Model 3	0.342	0.258	0.205	Model 3	0.268	0.277	0.285
	± 0.002	± 0.004	± 0.000		± 0.011	± 0.023	± 0.030

Table 3 Performance of the best models from each data set using fuzzy habitat preference models (FHPMs) optimized based on pairwise comparison (i) and FHPMs optimized based on means squared error (MSE) (ii), of which model 1 is the FHPMs developed using the data set of 14 October, model 2 is those of 5 November, model 3 is those of 9 November 2004. Each model was tested using three data sets.

(i) pairwise comparison-based optimization				(ii) MSE-based optimization			
Data set used				Data set used			
MSE	Oct. 14	Nov. 5	Nov. 9	MSE	Oct. 14	Nov. 5	Nov. 9
Model 1	0.097	0.090	0.260	Model 1	0.069	0.081	0.207
Model 2	0.287	0.080	0.169	Model 2	0.137	0.066	0.196
Model 3	0.418	0.166	0.177	Model 3	0.175	0.091	0.142
AUC	Oct. 14	Nov. 5	Nov. 9	AUC	Oct. 14	Nov. 5	Nov. 9
Model 1	0.835	0.799	0.870	Model 1	0.879	0.829	0.911
Model 2	0.727	0.838	0.906	Model 2	0.814	0.865	0.915
Model 3	0.703	0.815	0.876	Model 3	0.773	0.862	0.926
MP	Oct. 14	Nov. 5	Nov. 9	MP	Oct. 14	Nov. 5	Nov. 9
Model 1	0.228	0.281	0.304	Model 1	0.257	0.296	0.368
Model 2	0.339	0.235	0.226	Model 2	0.250	0.255	0.301
Model 3	0.340	0.249	0.205	Model 3	0.259	0.258	0.256

4 Discussion

The present results demonstrated the applicability and limitations of pairwise comparison in the optimization of an FHPM, from which some possible improvements of the present FHPM can be drawn. As we observed in Fig. 4, AUC and MSE could not be minimized by minimizing MP, and vice versa. The insensitive response of MP would be ascribed to its qualitiveness. An MP is a quantitative measure which evaluates the pairwise relationship between observed fish population density and composite habitat preference calculated by an FHPM. Due to the qualitiveness of MP, a GA could not optimize the singleton values, from which variable HPCs were produced. Since the present GA is a binary-coded scheme, the GA could not find an optimal solution due to discretized search space. The variance of HPCs can therefore be reduced by a real-coded GA. Fixed membership functions of the antecedent part of FHPMs could also be a cause of the insensitiveness, and thus learning habitat preference of the fish was restricted. This point could be improved by tuning the membership functions, which should, however, be limited to a certain extent in order to keep the semantics predefined from expert knowledge on the fish [15].

In contrast to the MP, the AUC positively responded to the MSE (Fig. 5). This could be because the FHPMs based on MSE could classify presence and absence of the fish by minimizing the errors between predicted and observed fish population density. For instance, an error at the point where no fish was observed but a model predicted fish presence is equal to a false positive in presence-absence analysis. Reducing the error can thus improve model performance with regard to AUC.

In addition to model performance, interpretation of an HPC is important in habitat modelling, from which users can learn how a target species would respond to a given condition through model output [4]. This is also important because it can

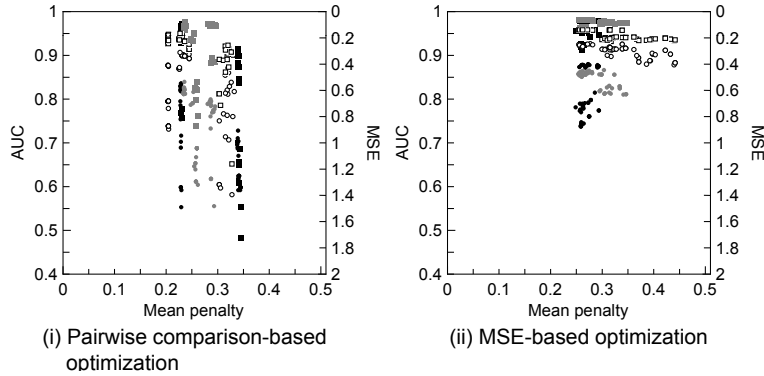


Fig. 4 Scatter diagrams between mean penalty (MP) and the area under receiver operating characteristics curve (AUC, circles) or mean squared error (MSE, squares): (i) fuzzy habitat preference models (FHPMs) optimized based on pairwise comparison and (ii) FHPMs optimized based on MSE. Black marks are the test results using the data set of 14 October, grey marks are those of 5 November, and white marks are those of 9 November 2004. Note that the axis for MSE is inverted.

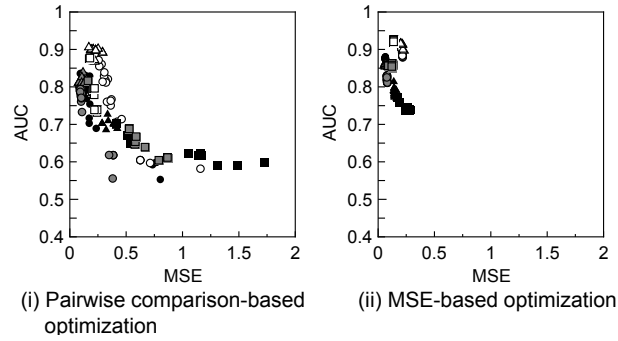


Fig. 5 Scatter diagrams between the area under receiver operating characteristics curve (AUC) and mean squared error (MSE): (i) fuzzy habitat preference models (FHPMs) optimized based on pairwise comparison and (ii) FHPMs optimized based on MSE. Circles are the results of FHPMs developed using the data set of 14 October, triangles are those of 5 November, and squares are those of 9 November. Black marks are the test results using the data set of 14 October, grey marks are those of 5 November, and white marks are those of 9 November 2004.

further be used for designing management plans and options for a target ecosystem. The interpretability of an HPC depends on the monotonicity toward a maximal preference point and the variance of HPCs obtained. The variance of HPCs would indicate uncertainty in fish habitat modelling. On the one hand, one of the variable HPCs may have a higher performance in a test data set. On the other hand, variable HPCs are not interpretable for users and may also deteriorate model performance. The variance observed in the HPCs of FHPMs based on pairwise comparison is ascribed to the algorithm used. In the pairwise comparison, only a qualitative relationship between modelled habitat preference and fish population density was considered, in which differences in degree of preference did not matter. Therefore, the variance in HPCs cannot be reduced by pairwise comparison if the shape is the same (even though the degrees of preference are different). In the present results, the effectiveness of pairwise comparison was not so prominent as we expected. This could be partially because the present FHPMs regulate the shape of an HPC during the optimization by a GA. If this is the case, pairwise comparison may work well when applied to habitat models without any regulation in the optimization process. Better results can be achieved by pairwise comparison using a set of individual habitat preferences for each habitat variable, instead of using composite habitat preferences as applied in this study. Different penalty assignments such as relaxation or rigidification of penalty criteria can also improve the model performance or the interpretability of the result. Further studies are necessary to clarify the mechanism of performance improvement as well as better algorithms for preference learning using pairwise comparisons, which contributes to the establishment of a sound methodology for habitat assessment.

Acknowledgements This study was partly supported by a Grant-in-aid for Young Scientists B from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan. W.W.

is supported as a postdoc by the Research Foundation of Flanders (FWO Vlaanderen). The authors thank Prof. E. Hüllermeier (Marburg University, Germany) whose comments helped formalize the concept of this study. S.F. thanks Prof. K. Hiramatsu (Kyushu University, Japan) for his assistance in the early phase of this study.

References

1. Adriaenssens V, De Baets B, Goethals P, De Pauw N (2004) Fuzzy rule-based models for decision support in ecosystem management. *Sci Total Environ* 319: 1–12.
2. Ahmadi-Nedushan B, St-Hilaire A, Bérubé B, Robichaud E, Thiémonge N, Bobée B (2006) A review of statistical methods for the evaluation of aquatic habitat suitability for instream flow assessment. *River Res Appl* 22: 503–523.
3. Bovee KD, Lamb BL, Bartholow JM, Stalnaker CB, Taylor J, Henriksen J (1998) Stream habitat analysis using the instream flow incremental methodology. U.S. Geological Survey, Biological Resources Division Information and Technology Report. USGS/BRD-1998-0004.
4. Elith J, Graham CH (2009) Do they? How do they? Why do they differ? On finding reasons for differing performances of species distribution models. *Ecography* 32: 66–77.
5. Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ Conserv* 24 (1): 38–39.
6. Fukuda S, De Baets B, Mouton AM, Waegeman W, Nakajima J, Mukai T, Hiramatsu K, Onikura N (2011) Effect of model formulation on the optimization of a genetic Takagi-Sugeno fuzzy system for fish habitat suitability evaluation. *Ecol Model* 222: 1401–1413.
7. Goldberg D (1989) Genetic algorithms in search, optimization, and machine learning. Addison-Wesley.
8. Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. *Ecol Model* 135: 147–186.
9. Hiramatsu K, Fukuda S, Shikasho S (2003) Mathematical modeling of habitat preference of Japanese medaka for instream water environment using fuzzy inference. *Trans JSIDRE* 228: 65–72 (in Japanese with English abstract).
10. Hüllermeier E, Fürnkranz J, Cheng W, Brinker K (2008) Label ranking by learning pairwise preferences. *Artif Intell* 172: 1897–1916.
11. Lechowicz MJ (1982) The sampling characteristics of electivity indices. *Oecologia (Berl.)* 52: 22–30.
12. Mouton AM, De Baets B, Goethals PLM (2010) Ecological relevance of performance criteria for species distribution models. *Ecol Model* 221: 1995–2002.
13. Pino-Mejías R, Cubiles-de-la-Vega MD, Anaya-Romero M, Pascual-Acosta A, Jordn-Lpez A, Bellinfante-Crocci N (2010) Predicting the potential habitat of oaks with data mining models and the R system. *Environ Modell Softw* 25: 826–836.
14. Takagi T, Sugeno M (1985) Fuzzy identification of systems and its applications to modelling and control. *IEEE Trans Systems Man Cybernet* 15: 116–132.
15. Van Broekhoven E, Adriaenssens V, De Baets B (2007) Interpretability-preserving genetic optimization of linguistic terms in fuzzy models for fuzzy ordered classification: An ecological case study. *Int J Approx Reasoning* 44: 65–90.